Final Technical Report

USGS/NEHRP Award No: 06HQGR0054

IMPROVING EARTHQUAKE LOCATIONS IN NORTHERN CALIFORNIA USING WAVEFORM-BASED DIFFERENTIAL TIME MEASUREMENTS

Principal Investigator: Felix Waldhauser

Co-PI: David Schaff

Lamont-Doherty Earth Observatory of Columbia University
Palisades, NY 10964
Tel: (845) 365 8538; Fax: (845) 365 8150
felixw@ldeo.columbia.edu

Project Period: 01/01/2006 – 12/31/2007

NEHRP Element III
Key Words: Seismology, Source Characteristics, Database

Research supported by the U.S. Geological Survey (USGS), Department of the Interior, under USGS award number 05HQGR0051. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Government.

Final Technical Report

USGS/NEHRP Award No: 06HQGR0054

IMPROVING EARTHQUAKE LOCATIONS IN NORTHERN CALIFORNIA USING WAVEFORM-BASED DIFFERENTIAL TIME MEASUREMENTS

Principal Investigator: Felix Waldhauser

Co-PI: David Schaff

Lamont-Doherty Earth Observatory of Columbia University
Palisades, NY 10964
Tel: (845) 365 8538; Fax: (845) 365 8150
felixw@ldeo.columbia.edu

ABSTRACT We simultaneously re-analyzed two decades (1984-2003) of the digital seismic archive of Northern California using waveform cross correlation (CC) and double-difference (DD) methods to improve the resolution in hypocenter locations in the existing earthquake catalog generated at the Northern California Seismic Network (NCSN) by up to three orders of magnitude. We used a combination of ~3 billion CC differential times measured from all correlated pairs of events that are separated by less than 5 km and ~7 million P-wave arrival time picks listed in the NCSN bulletin. The data is inverted for precise relative locations of 311,273 events using the DD method. The relocated catalog is able to image the fine-scale structure of seismicity associated with active faults, and reveals characteristic spatio-temporal structures such as streaks and repeating earthquakes. We find that 90% of the earthquakes have correlated P- and S-wave trains at common stations, and 12% are co-located repeating events. An analysis of the repeating events indicates that uncertainties at the 95% confidence level in the existing network locations are on average 0.7 km laterally and 2 km vertically. Correlation characteristics and relative location improvement are remarkably similar across most of Northern California, implying the general applicability of these techniques to image high-resolution seismicity caused by a variety of plate-tectonic and anthropogenic processes.

We have developed a real-time procedure (DD-RT) that rapidly relocates new earthquakes relative to nearby events in the new DD catalog. The DD-RT software currently runs on a test-bed at Lamont, using near real-time parametric and waveform data feeds from the NCSN and the NCEDC for new events, and a locally stored archive of seismic data for past events. We evaluated the performance of the new monitoring system by back-testing it with events that occurred in the past. This work demonstrates that consistent long-term seismic monitoring and data archiving practices, as followed at the NCSN and NCEDC, are key to increase resolution in existing hypocenter catalogs, and to estimate the precise location of future events on a routine basis.

1. Overview of Investigations

This final report covers the activities performed between January 1, 2006 (start date of the project) and December 31, 2007. The initial project was for one year, but has been extended by an additional year at no cost to the end of 2007 to accommodate some delays that we experienced during the second part of the project. The work described in this report is being undertaken by the principle investigator Felix Waldhauser and by co-PI David Schaff. The research includes the development of a final cross-correlation based double-difference earthquake catalog for events that occurred in Northern California between 1984-2003, an assessment of the improvement in the new hypocenter locations over existing locations produced at the Northern California Seismic Network (NCSN), and the development of a real-time double-difference location procedure that makes use of the new DD catalog and its associated waveform and phase pick archives. In this report we do not elaborate on the procedures we used to develop the new catalog, as we have described them in more detail in our last report for NEHRP project 05HQGR0051, of which this project is a continuation, and in a paper in press with JGR (Waldhauser and Schaff, 2008).

2. Investigations undertaken

Relocation of the NCSN catalog (1984-2003)

The new DD catalog (NCAeq_DD.1984-2003.v1.0; Waldhauser and Schaff, 2008) (Figure 1a) includes 311,273 earthquakes recorded at 6 or more stations for relocation, and is derived from a combination of differential travel times from ~7 million of NCSN P-wave picks, and 1.7 billion P-wave and 1.2 billion S-wave cross correlation differential times described in Schaff and Waldhauser (2005). Both pick and cross correlation differential times were combined in a double-difference inversion to insure location precision of correlated events to the accuracy of the cross correlation data, and of those that do not correlate to the accuracy of the pick data (Waldhauser and Ellsworth, 2000; Waldhauser, 2001). A detailed description of the relocation procedure, including a thorough investigation of the potential effects of un-modeled 3-D structures on our 1-D double-difference solutions, is given in Waldhauser and Schaff (2008) (see also Final Project Report of NEHRP grant 05HQGR0051).

The relocated catalog has a root mean square (RMS) of the weighted pick differential time residuals of 0.017 s, compared to 0.124 s before relocation. The weighted RMS of the cross-correlation data is 0.004 s after relocation. The new double-difference catalog reveals a focused view of the complex distribution of seismicity of northern California. Discrete faults are imaged with unprecedented detail in tectonic regions as diverse as the San Andreas Fault system (SAF), the Mendocino Triple Junction (MTJ), the Long Valley Caldera (LVC), and the region of induced earthquakes at the Geysers Geothermal Field (GGF) (Figure 2). In addition, artifacts in the NCSN locations (arrow in Figure 2), most likely caused by effects related to transitions in regional 1D models used for routine locations, are removed.

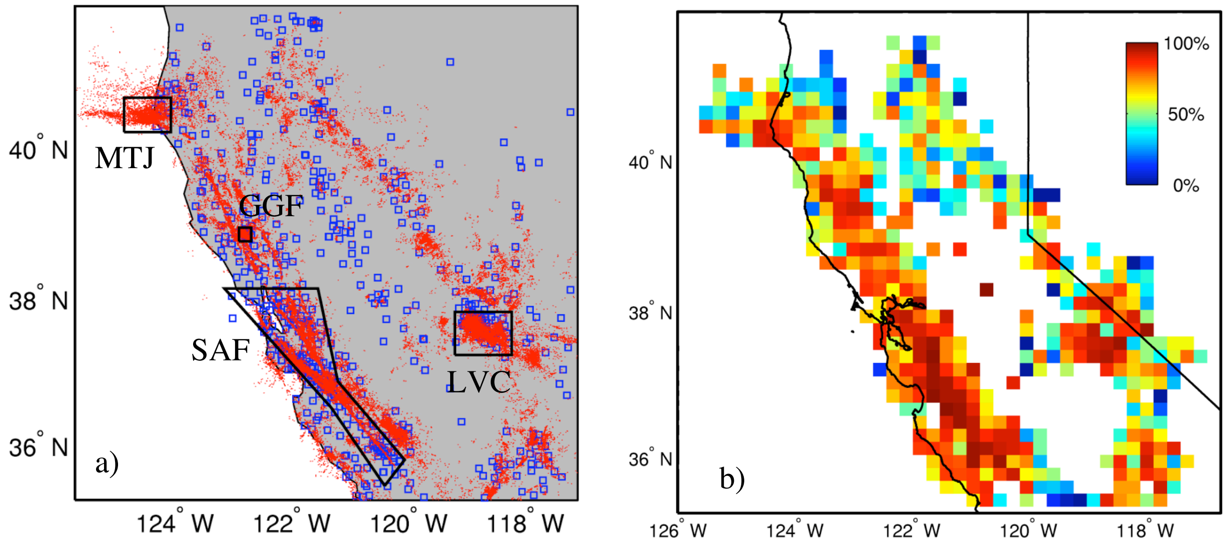

Figure 1 a) 311,273 relocated earthquakes (gray dots) in Northern California recorded by 6 or more stations (squares) of the Northern California Seismic Network (NCSN) between January1984 – May 2003. Labeled polygons show the four focus regions discussed in this study: SAF: selected area of the San Andreas Fault system, including the San Andreas, Hayward, and Calaveras faults; MTJ: Mendocino Triple Junction; LVC: Long Valley Caldera; GGF: Geysers Geothermal Field. b) Percentage of correlated earthquakes across northern California, displayed within cells of 20 by 20 km. Only cells with 10 or more events are shown. Black lines denote coast and state line.

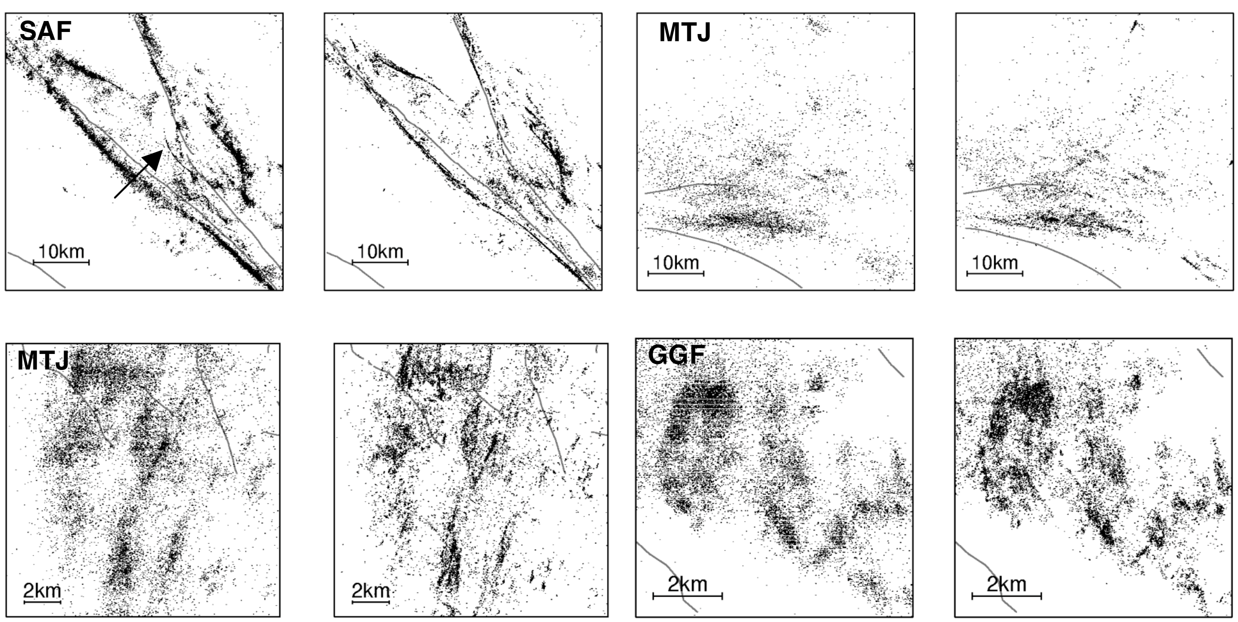

Figure 2 Comparison between network and double-difference locations. Map views of representative areas in each of the four focus regions are shown. Note the networks of discreet faults imaged in the tectonic regions SAF, MTJ, and LVC, compared to the 'clouds' imaged in the region of induced seismicity at GGF. Lines indicate traces of faults mapped at the surface.

We find that 90% (or 185,601 events) of all earthquakes with digital waveforms available from the NCEDC correlate (Figure 1b). We define an earthquake pair as correlated when at least four first-arriving P-wave trains are similar at a cross-correlation coefficient of 0.7 or greater in the frequency band 1.5-15 Hz. Similar percentage values of 94% and 87% are found when we require at least 3 and 5 similar P-waves trains, respectively. Correlated earthquakes occur widespread across northern California with high concentrations being observed in a variety of tectonic settings that include predominantly strike-slip transform faults (SAF, MTJ), subduction zones (MTJ), volcanic areas (LVC), and geothermal fields (GGF) (Figure 1b; Table 1).

The decrease of the P-wave correlation coefficients for correlated event pairs with increasing hypocentral separation is remarkably similar within the four regions, dropping from an average of *Cf*~0.9 for nearby hypocenters to ~0.78 for events separated by 5 km (Figure 3a). The highest decay rate is observed for event pairs in the SAF region, where the abundance of streaks and repeating events along the San Andreas and Calaveras faults (Rubin et al., 1999; Schaff et al., 2002; Waldhauser et al., 2004) produces high *Cf* values at short separation distances, and the strong and complex structural variations caused by the interaction of the Pacific and American plate result in a rapid decay of *Cf* with increasing event separations. Coefficients are consistently lowest at the GGF where the events are induced by geothermal production activities, fracturing undisturbed porous rock along new faults whose orientations are random and varies rapidly over short distances (Oppenheimer, 1986). In addition, GGF may exhibit time-dependent short-wavelength velocity variations due to the movement of fluids. As a result, the majority of earthquakes at GGF correlate over distances less than 1 km (Figure 3b). In comparison, correlation measurements in the three tectonic regions SAF, LVC, and MTJ are obtained over a broad range of separation distances between 0 and 5 km (Figure 3b).

S-waves show *Cf* values that are overall lower and break down faster with increasing hypocenter separation compared to the P-wave coefficients because of their shorter wavelengths due to the slower wavespeed (Figure 3c). S-waves typically correlate on pairs separated by less than 2 km, regardless of the type of tectonic region in which they occur (Figure 3d). S-waves also tend to de-correlate faster than P-waves due to their contamination with dissimilar P-wave coda.

Table 1. Statistics of the cross-correlation based double-difference catalog. DD relative location errors are median and mean of the major axes of the horizontal and vertical projection of the 95% confidence ellipsoids calculated from 200 bootstrap samples for each event. Network relative location errors are estimated from an analysis of repeating events. All errors in km.

	ALL	SAF	LVC	MTJ	GGF
Number of relocated events (% with waveforms)	311,273 (66)	81,679 (82)	115,751 (53)	7,636 (85)	46,448 (53)
% correlated events	90	96	95	88	92
% repeating events	11.9	27.1	4.9	2	5.2
Median DD errors (x/z):	0.050 / 0.047	0.039 / 0.030	0.043 / 0.043	0.196 / 0.070	0.041 / 0.045
Mean DD errors (x/z):	0.450 / 0.290	0.308 / 0.068	0.199 / 0.103	1.025 / 0.227	0.090 / 0.071
Network errors (x/z)					
95% confidence level	0.715 / 2.069	0.749 / 1.675	0.513 / 1.662	1.552 / 2.361	0.280 / 1.818
mean	0.172 / 0.257	0.158 / 0.243	0.138 / 0.208	0.340 / 0.325	0.096 / 0.190
median	0.111 / 0.160	0.108 / 0.160	0.107 / 0.140	0.300 / 0.220	0.079 / 0.130
maximum	7.983 / 6.958	3.670 / 6.958	2.048 / 3.218	2.242 / 2.600	0.887 / 1.947

Correlation coefficients of both P- and S-waves decrease linearly with increasing difference between an event pairs' magnitudes, with most correlations obtained for pairs with a magnitude difference less than 2.

Our results show that the ability for two events to produce similar seismograms (in the frequency band 1.5-15 Hz), from which we can precisely measure phase delay times at common stations, primarily depends on the distance between their hypocenters and the difference between their magnitudes, and less so on the tectonic environment in which the events occur. This indicates that most of the seismicity in northern California occurs along repeatedly breaking faults that are sufficiently smooth and long to generate earthquakes with similar seismograms over long separation distances.

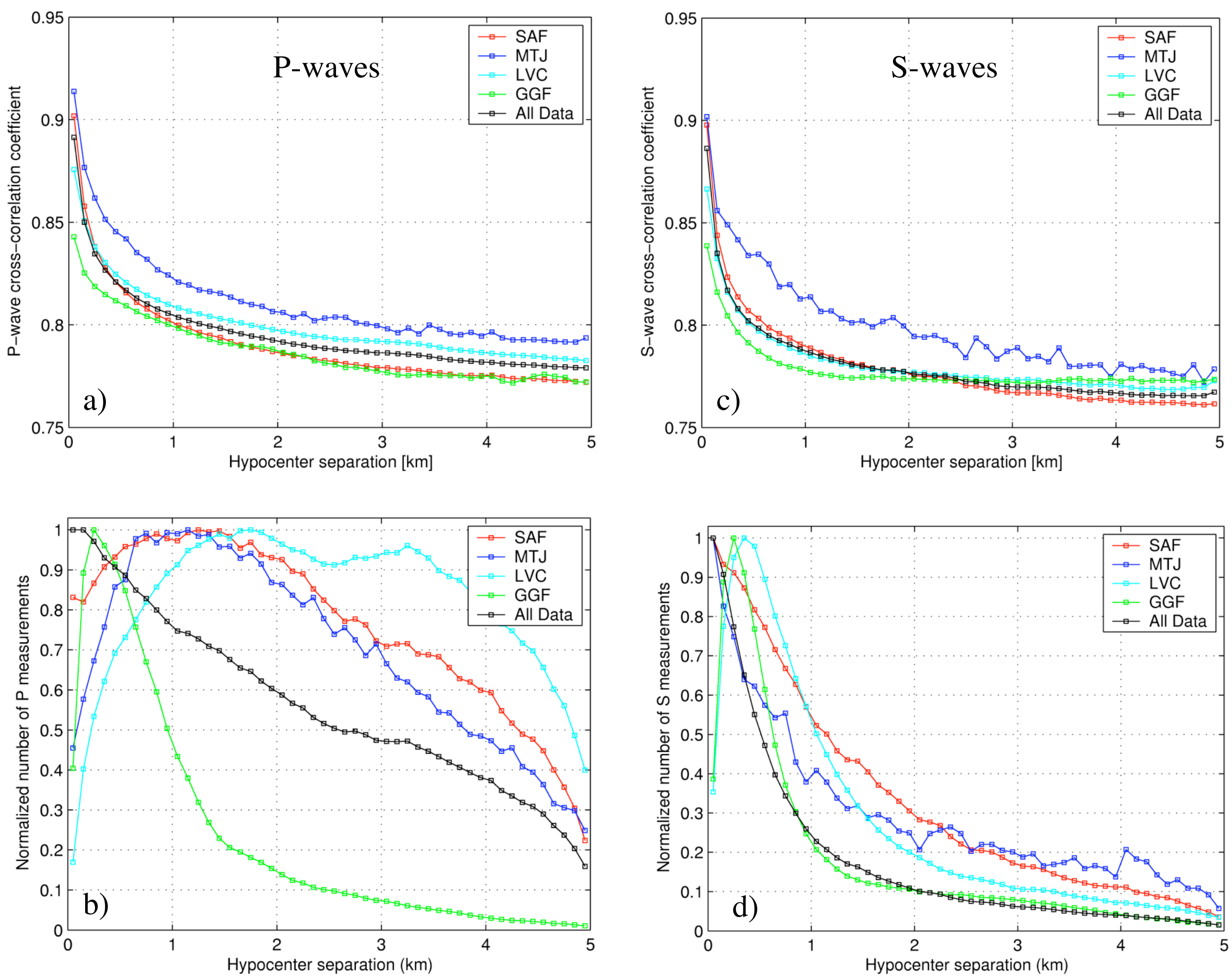

Figure 3 Distribution of P- and S-wave cross-correlation coefficients (a,c) and normalized number of correlations (b,d) shown as means within bins of 0.1 km distance between correlated events. P-wave data (a,b) are based on 152 million P-wave cross-correlation coefficients with *Cf* ≥ 0.7 measured from 23 million pairs of correlated earthquakes, S-wave data (c,d) are from ~44 million S-wave cross-correlation coefficients (*Cf* ≥ 0.7) derived from ~7 million pairs of correlated events. Statistics are shown for all correlated events in northern CA (open circles) and for individual tectonic regions SAF, MTJ, LVC, and GGF. See Figure 1a for abbreviations and geographic locations.

Evaluation of location improvement

The precision of the cross-correlation measurements and the improvement over existing picks is most readily assessed by using repeating earthquakes — a special category of correlated earthquakes that rupture the same fault patch more than once and therefore exhibit highly correlated waveforms and virtually zero delay times (Poupinet et al., 1984; Vidale et al., 1994; Nadeau et al., 1995). Repeating events have been shown to exist predominantly in the creeping sections of the San Andreas fault system. We search for repeating events in the double-difference catalog by selecting all pairs of events that produce P-wave trains with a mean $Cf \geq 0.9$ at 5 or more common stations, have well constrained hypocentral separations that are smaller than their respective rupture radius calculated from a circular, 3MPa stress drop model, and have similar magnitudes ($\pm M_L 0.3$) (Waldhauser and Ellsworth, 2002). We find a total of 24,438 repeating events that represent 12% of all events with waveforms. They occur in 7,406 clusters of between 2 and 33 events with magnitudes up to M_L 4.3 throughout northern California (Table 1). While sequences with at least 2 repeating events are widespread, sequences with at least 5 events concentrate in the four regions SAF, MTJ, LVC, GGF. We find sequences with at least 10 events only on the creeping section of the San Andreas and Calaveras faults where they appear to image the transfer of fault creep at seismogenic depths from the San Andreas to the Calaveras fault.

The median of the absolute cross-correlation differential times of these repeating events, after subtracting the mean in each cluster, is 0.002 s for both P- and S-waves, and the standard deviation (SD) is 0.01 s. In comparison, the median of the corresponding absolute differential times formed from the P-wave picks is 0.023 s (SD = 0.15 s), which is ~14 times less precise than the cross correlation data for repeating events. These metrics are derived from the original measurements before relocation and therefore include the outliers that form the long tails of the distributions, especially in the cross correlation data. These outliers are easily detected by their large residuals and typically down-weighted or removed during the double-difference inversions. The precision of the delay time measurements decreases with increasing hypocenter separation, as waveforms become more dissimilar due to changes in the focal mechanisms and differences in the ray paths (e.g. Waldhauser and Ellsworth, 2000; Schaff et al., 2004).

Since we locate highly correlated repeating events to the precision of several meters to a few tens of meter (Waldhauser and Ellsworth, 2000; Rubin, 2002), the deviations of the corresponding network locations from the centroid location of each group of repeating events reflect their relative location error. We find that these network locations have errors at the 95% confidence level of 0.7 km horizontally and 2 km vertically, and maximum mislocations of 8 km and 7 km, respectively (Table 1). Cross-correlation based DD locations for repeating events in well-monitored regions thus represent a relative location improvement of up to a factor of ~1000 over existing network locations. The greater improvement in vertical control is due to the additional S-wave differential times obtained via cross-correlation. Both network and double-difference relative location errors are largest for events near MTJ, and smallest for those in the GGF region (Table 1), reflecting differences in availability and coverage of seismic stations.

Double-difference locations of the repeating events based on phase picks alone (i.e., only minimizing model errors in the network locations but not reducing pick uncertainty) have errors at the 95% confidence level of 0.17 km horizontally and 0.7 km vertically, indicating a factor of ~4 improvement in location precision over existing network locations. The significant improvement obtained by applying double-differences to picks alone is also visually demonstrated. Pick based DD locations are closer to the CC based DD locations than they are to the catalog locations, imaging detailed fault structures at the scale of a few hundreds of meters.

Real-time double-difference event relocation

We initially proposed to assist Dr. David Oppenheimer and Dr. Jim Luetgert of the USGS with developing and implementing a real-time DD procedure in Menlo Park. However, it has proven impossible to work efficiently via a VPN connection between computers at Lamont and the USGS. Therefore, we decided to develop the whole real-time double-difference location software (DD-RT) on a test-bed here at Lamont, using near-real time data feeds from the USGS in Menlo Park (phase picks and initial locations via email) and the Northern California Earthquake Data Center (NCEDC) at the Berkeley Seismological Laboratory (waveforms via Doug Neuhauser's *swc* software). The core of the DD-RT process is a real-time version of the hypoDD software package (Waldhauser, 2001), with numerous additional programs for collecting, selecting, and processing the data, controlling the process flow, and generating a dynamically updated web site for displaying and accessing the relocation results over the internet. We are in the process of testing a beta version of the DD-RT software that is currently running her at Lamont, and which we plan to migrate to a dedicated computer in Menlo Park sometime later this year. In the following we briefly describe the basics of the procedure and show some results from back-testing it.

The DD-RT procedure relocates new events relative to the high-resolution double-difference catalog described above by computing delay times between a new event and its neighbors and inverting these data using a real-time version, *hypoDD-RT*, of the double-difference algorithm *hypoDD* described in Waldhauser and Ellsworth (2000) and Waldhauser (2001). Parametric data in Hypoinverse archive format (including phase picks, hypocenter location and origin time, and magntitude estimates) is automatically sent by the USGS via email after a new event has been processed at the NCSN (David Oppenheimer, pers. communication). Seismograms of new events, two minutes long and starting 10 s before the origin time, are automatically requested from the NCEDC using the program *swc* (Simple Waveform Client; Doug Neuhauser, pers. communication). The *swc* program access the Simple Wave Server (SWS) at the NCEDC which stores the real-time DART waveforms, and downloads the seismograms in miniSEED format to the local disk.

After we received the parametric data of a new event, the DD-RT procedure finds the best neighboring reference events in DD base catalog relative to which the new event is going to be relocated. For these reference events we find the corresponding pick and waveform data from a locally stored archive of some 7 million phase picks and 15 million seismograms. We form differential times between the new event and its neighboring reference events from the picks directly and from the waveforms via cross correlation, and invert the combined data set for the vectors connecting the new event to its reference events. The DD-RT solutions are automatically being posted to a web site (www.ldeo.columbia.edu/~felixw/DDrtCISN) from which they can be openly accessed. Access to the DD-RT web-site is currently password-protected while beta-testing is ongoing, but will be openly accessible as soon as these tests are completed.

We back-tested the DD-RT procedure by treating past events as new events and relocate them relative to the DD base catalog, excluding the archived event to be relocated. The DD-RT locations are then evaluated in terms of their deviation from the corresponding location in the DD base catalog. We use 2,360 events that occurred in 2003, randomly selected within cells of 20x20 km across northern California to achieve uniform aerial coverage (see Figure 1b for cell locations). A summary of the differences between the DD-RT locations and the corresponding locations in the DD base catalog is given in Figure 4. The most precise relocations are achieved along the San Andreas fault at Long Valley, where epicentral differences between the DD-RT locations and the corresponding locations in the DD base catalog have medians typically less than 100 m, and less than 50 m in well monitored regions (Figure 4c, top panel). Differences in depths are somewhat larger (< 500 m), except for the well monitored regions along the San Andreas fault where median differences are smaller than 100 m (Figure 4c, bottom panel). Pick based DD-RT relocations are typically within 100 m horizontally and 500 m vertically along the San Andreas fault and Long Valley, and within 500 m horizontally and 1000 m vertically elsewhere. For 12% (3%) of the seismically active area we achieve horizontal (vertical) location improvement of a factor of 10 or better, for 40% (20%) a factor of 5 or better, and for 80% (54%) of a factor of 2 or better. Epicenter locations improve in all areas, while hypocenter locations appear to get worse in 5% of the area.

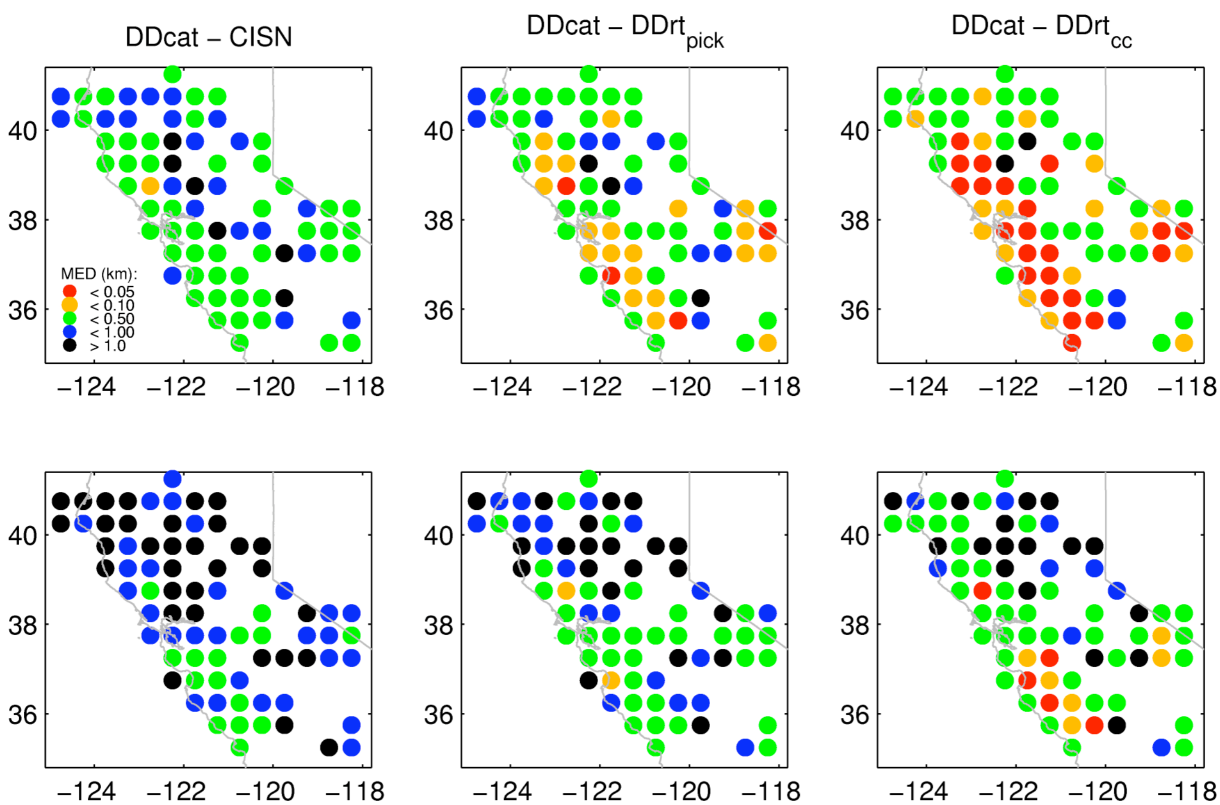

Figure 4 Results from back-testing the DD-RT procedure using 2360 events from 2002. Differences between epicenter (top panels) and depth (bottom panels) locations in the DD catalog and corresponding locations in the NCSN catalog (left), the pick-based DD-RT locations (middle panels), and the cross-correlation based DD-RT locations (right panels). Median differences are shown for events in 20 x 20 km cells.

4. References

Nadeau, R. M., W. Foxall, T. V. McEvilly (1995), Clustering and periodic recurrence of microearthquakes on the San Andreas Fault at Parkfield, *Science, 267*, 503-507.

Oppenheimer, D. (1986), Extentional tectonics at the Geysers Geothermal Area, California, *J. Geophys. Res., 91*, 11,463.

Poupinet, G.,W.L. Ellsworth, and J. Fréchet (1984), Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras fault, California, *J. Geophys. Res*., *89*, 5719-5731.

Rubin, A. M., D. Gillard, J.-L. Got (1999), Streaks of microearthquakes along creeping faults, *Nature, 400*, 635-641.

Rubin, A. (2002), Using repeating earthquakes to correct high-precision earthquake catalogs for time-dependent station delays, *Bull. Seism. Soc. Am. 92*, 1647-1659.

Schaff, D.P., G. Bokelmann, G.C. Beroza, F. Waldhauser, and W.L. Ellsworth (2002), High-resolution image of Calaveras Fault seismicity, *J. Geophys. Res., 107*, 2186, doi: 10.1029/2001JB000633.

Schaff, D. P., G. Bokelmann, W. L. Ellsworth, E. Zanzerkia, F. Waldhauser, and G.C. Beroza (2004), Optimizing correlation techniques for improved earthquake location, *Bull. Seism. Soc. Am.*, 94, 705-721.

Schaff, D.P. and F. Waldhauser (2005), Waveform cross-correlation-based differential travel-time measurements at the Northern California Seismic Network, *Bull. Seism. Soc. Am.*, *95*, 2446-2461.

Vidale, J. E., W. L. Ellsworth, A. Cole, C. Marone (1994), Variations in rupture process with recurrence interval in a repeated small earthquake, *Nature, 368*, 624-626.

Waldhauser, F. and W. L. Ellsworth (2000), A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, *Bull. Seism. Soc. Am.*, 90, 1353-1368.

Waldhauser, F. (2001), HypoDD: A computer program to compute double-difference hypocenter locations, *U.S. Geol. Surv. open-file report*, 01-113, Menlo Park, California.

Waldhauser, F., W.L. Ellsworth, D.P. Schaff, and A. Cole (2004), Streaks, multiplets, and holes: High-resolution spatio-temporal behavior of Parkfield seismicity, *Geophys. Res. Lett. 31*, L18608, doi:10.1029/2004GL020649.

Waldhauser, F. and D.P. Schaff (2008), Large-scale relocation of two decades of northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.,* in press.

5. Reports published related to this project

Peer-reviewed journal publications:

Waldhauser, F. and D.P. Schaff (2008), Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.,* in press.

Thurber, C., H. Zhang, F. Waldhauser, J. Hardebeck, A. Michael, and D. Eberhart-Phillips, Three-Dimensional Compressional Wavespeed Model, Earthquake Relocations, and Focal Mechanisms for the Parkfield, California, Region, *Bull. Seism. Soc. Am. 96*, 38-49, 2006.

Conference abstracts (*invited):

Waldhauser, F. and D. P. Schaff, A cross-correlation double-difference catalog for northern California and its use in near real-time event relocation, 5th Annual Northern California Earthquake Hazards Workshop, Menlo Park, CA, January 23-24, 2008.

Waldhauser, F., D. P. Schaff, Back to the Future: Long-Term Seismic Archives Revisited, *Eos Trans. AGU, 88*(52), Fall Meet. Suppl., Abstract U23C-1447, 2007.

Waldhauser, F. and D. Schaff, Improving earthquake locations in northern California using waveform-based differential-time measurements, 4th Annual Northern California Earthquake Hazards Workshop, Menlo Park, CA, January 18-19, 2007.

*Waldhauser, F., and D. Schaff, Massive-scale Waveform-based Double-difference Event Relocation in Northern California, *Eos Trans. AGU,* 87(36), West. Pac. Geophys. Meet. Suppl., Abstract S14B-02, 2006.

*Waldhauser,F., H. Abend, D. Bohnenstiehl, W. Kim, P. Richards, D. Schaff, M. Tolstoy, Fine-scale Seismicity of Earth's Interior: Regional- and Global-scale Double-difference Applications to Study Plate-tectonic Processes, SSA Annual Meeting, San Francisco, *Seism. Res. Lett.*, 77 (2), 2006.

Thurber, C., H. Zhang, F. Waldhauser, A. Michael, J. Hardebeck, D. Eberhart-Phillips, Three-Dimensional Compressional Wavespeed Model, Earthquake Relocations (1966- 2005), and Focal Mechanisms for the Parkfield, California, Region, *Eos Trans. AGU, 87*(52), Fall Meet. Suppl., Abstract S32A-01, 2006.

6. Available Data and Products

The new double-difference catalog developed under this grant, NCAeq_DD.1983-2003.v1.0, is openly available and can be requested from the PI. We are working on setting up a website that will host the catalog, and tools to investigate the data online before downloading.

Contact: Dr. Felix Waldhauser
Lamont-Doherty Earth Observatory
61 Route 9W
Palisades, NY 10964
845 365 8538
felixw@ldeo.columbia.edu

7. Final Remarks

We are in the process of updating the DD catalog, with events from June 2003 – December 2007, under a currently active project funded by the NEHRP program. The new updated DD catalog will then replace the existing DD catalog in the real-time double-difference procedure. Starting from January 2009 we expect to be able to maintain a continous DD catalog for northern California in near-real time.